

Chomsky と人工知能

ー 普遍文法とディープラーニングをめぐる論考 ー

石 崎 貴 士

(山形大学大学院教育実践研究科)

Critical Remarks on the Relationship between Universal Grammar and Artificial Intelligence

Takashi ISHIZAKI

This paper intends to make some critical remarks on the relationship between universal grammar (UG), which is an essential concept for language acquisition in linguistics, and artificial intelligence (AI), which, along with deep learning, is derived from neural network models in cognitive psychology. These two concepts are widely referred to, and sometimes regarded as interactive or connected even by experts in the respective fields. For example, one of the authorities of AI, Matsuo (2015), contends that AI should include UG. Although Chomsky, the founder of UG, evoked the cognitive revolution in linguistics and psychology in the 1960s, which is closely related to the progress of AI, the goals of the studies in UG and AI are completely different and in fact, contradictory. This difference is significant, and sometimes causes considerable controversy. This paper describes this difference in detail in order to clarify the issues to be discussed in further studies.

[キーワード] Chomsky, 普遍文法, 人工知能, ディープラーニング

1. はじめに

昨今の人工知能の発展は目覚ましい。チェス、将棋、囲碁の対局でプロを打ち負かしたばかりか、作曲、作画や小説執筆などクリエイティブな分野にまで人工知能は活躍の場を広げている。このままでは人工知能が人間の知を乗り越えてしまうのではないか—この「シンギュラリティ (技術的特異点)」をめぐる問題が現実味をもって語られ始めている¹⁾。そのような中、日本における人工知能ブームの火付け役ともなった人工知能研究の第一人者である松尾豊氏の『人工知能は人間を超えるか』(2015)には、Noam Chomsky の普遍文法に関する以下のような記述がある。

人は言葉を話す。特に、「文法」を使って文の形でものごとを描写したり、書き綴ったりする。では、文法はどのように獲得できるのだろうか。有名な言語学者のノーム・チョムスキー

氏は、人間は生得的な文法 (普遍文法) を備えていると言った。私の考えもこれに近い。(松尾 2015: 193)

それ (普遍文法) をコンピュータに埋め込まないと、人間と同じような文法を獲得するのは難しいかもしれない。(ibid., 195, 但し括弧内は筆者が添付)

しかしながら、Chomsky の主張する普遍文法は生得的にプログラミングされた制約を前提とする直列処理的な発想であるため、人工知能を支えるディープラーニングのような並列分散型の処理とは相容れないようにも思える。確かに、その起源を辿れば人工知能研究と Chomsky とは無縁ではない。むしろ、Chomsky によって当該領域に歴史的な転回がもたらされたとも言える。しかし、それでもなおこの違和感は拭えない。本論では、その理由を考究していく。

2. 認知科学の成立

1950年代後半は、言語学と心理学における一大転換期であり、そのいずれにおいても Chomsky が火付け役であった。言語学においては、それまで Leonard Bloomfield を旗振り役とする構造主義が主流であった (Bloomfield 1933)。これは言語学に初めて科学的な分析手法が導入されたものであり、もともとは文化人類学の分野で、文字を持たないアメリカ先住民の言語を研究する際に用いられた手法を応用したものである。それは採取した言語データをできる限り細かく記述していくというもので、直接構成素分析 (immediate constituent analysis, IC 分析) と呼ばれる。この手法から音声の最小単位である音素 (phoneme) や語の最小単位である形態素 (morpheme) が研究の対象とされるようになった。構造主義の言語学者たちは、言語の詳細な記述を通して、これらの最小単位には言語それぞれに固有の配列パターンがあることを見出し、それらを対立という概念で捉え体系づけることによって言語の本質に迫ろうとした。

ところが Chomsky は、その著書 *Syntactic Structures* の中で構造主義の言語学に反論する (Chomsky 1957)。以下の2文は、それを説明する好例として頻繁に参照される Chomsky (1964) からの引用である (Chomsky 1964: 34)。

- 1) John is easy to please.
(ジョンは喜ばせるのが簡単だ)
- 2) John is eager to please.
(ジョンは喜ばせるのに熱心だ)
《和訳は筆者が添付》

これらの2文は、表面上は全く同じ構造をしている。しかし、主語の John が 1) の文では不定詞 to please の意味上の目的語になっているのに対し、2) の文では不定詞 to please の意味上の主語になっている。なぜ、このような意味の違いが生じてしまうのか。この疑問については、IC 分析のような従来の構造主義的なアプローチでは説明ができない。Chomsky によるこの指摘によって構造主義言語学における致命的な欠陥が明らかになった。

一方、心理学においては B. F. Skinner が牽引したオペラント条件付けによる行動主義が当時の

主流であった。Ivan Pavlov による古典的条件付けでは、刺激と反応の連合における刺激を人為的にすり替えることによって行動 (としての反応) をコントロールしようとするが、この方法は人間による行動のように、もはや刺激の特定が困難であるような場合には適用できない。Skinner は、このような「刺激を特定できない行動 (反応)」を「オペラント (operant)」と名付け、オペラントに対しては、その後に「賞」や「罰」を与えることによってコントロールが可能であると主張した。これが「オペラント条件付け」である (林 1984)。

Skinner は、この条件付けの過程を「強化 (reinforcement)」, 賞や罰を「強化子 (reinforcer)」と呼び、その著書 *Verbal Behavior* の中で人間の言語行動についても強化によってコントロールが可能なオペラント (verbal operants, 言語オペラント) であるという考えを示した (Skinner 1957)。

これに対し Chomsky は、この *Verbal Behavior* についての書評の中で母語習得の観点から批判を展開している (Chomsky 1959)。Skinner の主張するような注意深く構成された言語強化のプログラムは、よほどアカデミックな家庭でなければ実行が不可能であろう。しかし、実際には上流階級の子どもだろうが移民の子どもだろうが、おかれていた社会的環境に関係なく子どもは周りで話されていることばから、同じように完全に同じように速く母語を習得していく。この母語習得の一律性は、Skinner の主張する「強化」のような経験論では説明できない。Chomsky は、このようなことができるのは子どもの脳に生まれつき文法能力が備わっているからだとして生得論を主張した。

Chomsky によって提示された言語学や心理学におけるこれらの問題を克服するためには、いずれの領域においても、音素や形態素といった表層的な要素ではなく、また経験という外的な要因でもない、より内的な深層へと研究の矛先を転回しなければならなかった。このことが認知科学を成立させる一因となった。

3. 計算主義と GB 理論

同時期に人工知能研究も、後に第一次人工知能ブームと呼ばれる最初の黄金期を迎えている。認知科学においても、この影響を受けて、いわゆる「計算主義」が標榜される。人間をコンピュータ

のような一種の情報処理体とみなす考え方である。そして、この傾向は先述の Chomsky の主張にも見ることができる。

母語習得に対する Chomsky の主張の背景には、経験論批判の他に、それを発展させた「刺激の貧困」という問題がある (Chomsky 1965)。母語習得においては、隔離児や障碍児など特殊な場合を除き、全ての子どもが生まれてから数年の間に母語の複雑な文法規則の体系をマスターする。しかし、そのような短期間に子どもが実際に耳にする言語の例は、質的にも量的にもとても十分とは言えない貧弱なものであるはずである。しかも、子どもは、これまで聞いたことのない文を発話することもできることから、単に耳にした形式を記憶しているのではなく、規則を習得していることは明らかである。この不思議な事象を説明するために Chomsky は、子どもの頭の中に予め生得的に文法体系のもととなるような規則「普遍文法」が備わっているからだと主張する。そして、この普遍文法を説明するのに Chomsky は、人間の脳をコンピュータに見立てたメタファーを用いている。Chomsky の主張するこのモデルは GB 理論 (統率・束縛理論, Government-Binding Theory) と呼ばれる (Chomsky 1981, 1988)。

まず、人間の脳内には「言語習得装置」なるものが存在し、「普遍文法」はその中核を担う。その基本的なスペックは、人間という種を共有する限りにおいて遺伝的に統一されており、普遍文法は全ての言語に通用する普遍的な「原理」と、その原理の枠内で個々の言語の特性を同一水準のオン・オフで決定していく項目別のスイッチのような「パラメータ」によって構成されている。各パラメータのデフォルトの値は遺伝的に決まっているが、その後、実際に耳にする言語データをもとにそれぞれでセッティングが行われ、個々の言語における文法が完成する。

例えば、そのようなパラメータの 1 つに有名な pro-drop がある。このパラメータは、文中の主語を省略することが文法的に容認可能であることを決定するものであり、これをオンに設定すると日本語のように主語が省略できる言語になり、オフにすると英語のように主語が省略できない言語になる²⁾。

また Chomsky は、このモデルの正当性を主張する論拠として、先述の「刺激の貧困」の他に、子

どもが母語を「肯定証拠 (positive evidence)」のみで習得することも挙げている³⁾。この場合、肯定証拠とは文法規則に適った言語の正しい用例のことを指している。人が規則を知るということは「何をしてよいのか」だけでなく「何をしてはいけないのか」についても知ることを意味する。通常、何をしてよいのかについては肯定証拠のみで学習が可能であるが、何をしてはいけないのかについて知るには肯定証拠だけでは不十分であり、「否定証拠 (negative evidence)」が必要となる。ここで言う否定証拠には「直接的な否定証拠」と「間接的な否定証拠」がある。直接的な否定証拠は、子どもが文法的に誤った発話をした際、その発話が誤りであることを大人から指摘・訂正されることを指し、間接的な否定証拠は、文法的に誤ったそのような用例を耳にしないという事実を指す。

本来であれば、規則を学ぶには直接的な否定証拠の方が有効であると言える。なぜなら、間接的な否定証拠については、そのような用例をたまたま耳にしていなかったかもしれないという可能性を排除できないからである。しかし、もし直接的な否定証拠が母語習得に関与していたら、それを経験した子どもと経験しなかった子どもとの間に差異が生じてしまう。それでは母語習得の一律性を説明できなくなってしまう。それ故、習得に関与しているのは間接的な否定証拠のみということになる。このように間接的な否定証拠のみで母語習得が成り立つのは、普遍文法が生得的に頭の中に備わっていて、それが制約としての機能を果たしているからだ Chomsky は主張している。

4. ニューラルネットワーク：研究の推移

Chomsky が人間の脳をコンピュータに見立てたのに対し、ニューラルネットワーク (後に人工知能研究の主領域となる) はコンピュータで人間の脳の働きを再現しようとしている。つまり、コンピュータ上に人間の脳内で構成されている神経細胞 (ニューロン) のネットワークを模したモデルを築き、それに学習をさせるのである。神経細胞は、シナプスによって他の神経細胞とつながっている。シナプスでは伝達効率に応じて電気的な刺激が伝わり、いくつもの他の細胞から伝わってきた刺激が当該の細胞の中で、あるボルテージ (閾値) にまで達すると当該の細胞自体が発火し、そ

の興奮が軸索を通過して、また別の細胞へと伝えられる。しかも、このシナプスでの伝達効率は絶えず変化する。これをシナプスの可塑性と言う。ニューラルネットワークにおけるシミュレーションでは、このような伝達効率を結合強度とみなし、ニューラルネットワークにおける学習とは、この結合強度の調整を意味する。以下、その歴史的な研究の推移について概観する⁴⁾。

本格的なニューラルネットワーク研究は、1958年にFrank Rosenblattが発表したパーセプトロンに始まる。パーセプトロンは、入力層と出力層のみのシンプルなニューラルネットワークである。この入力層と出力層の間の結合強度の調整は計算によって行われている。具体的には、出力層の各神経細胞のユニットで生じた実際の出力値と正解である教師信号とのズレの程度をエラー（誤差）として算出し、そのズレの程度に応じて神経細胞間の結合強度の調整を行っていく。パーセプトロンは、そのようなシンプルなネットワークであったが、学習や予測ができることから注目を集め、これにより人工知能研究は第一次ブームを迎える。しかし、1969年Marvin MinskyとSeymour Papertによってパーセプトロンが排他的論理和（XOR）のような線形分離不可能なパターンを識別できないことが示されると、このことが議論となり第一次人工知能ブームは終わりを告げた。

しかし、その後の研究によりパーセプトロンに中間層を入れて多層化すれば、XORのような非線形課題も解決できることがわかった。さらに、1986年David RumelhartとJames McClellandたちによって「誤差逆伝播法（backpropagation）」という、この中間層を持つ多層構造のパーセプトロンの結合強度を調整するための学習アルゴリズムが考案されると、人工知能研究は第二次ブームを迎えた。

神経細胞間の結合強度を調整すると、それに伴ってエラーの程度も変化していく。その軌跡をグラフにしたのがエラー曲線である。その調整の様子を、複雑なエラー曲線の上をビー玉が転がっていくイメージで捉えると、一番低い位置にある窪みにはまったところが、もっともエラーの少ないゴールとみなすことができる。そして、この窪みの位置は、その地点での曲線の傾斜（傾き）が0となることから曲線の導関数によって算出することができる。この「勾配降下（gradient descent）」を利用した学習アルゴリズムが誤差逆伝播法であ

る⁵⁾。

後は、この中間層の階層を増やしていけば、人間の脳細胞のネットワークに近づくことができるのではないか、そう考えられたが、実際には中間層の階層を増やしても満足な結果を得ることはできなかった。実は、多層構造のパーセプトロンは中間層の階層を増やしていくと、末端に行くほど誤差逆伝播法による調整が届かなくなってしまうのである。この「勾配消失（vanishing gradient）」の問題が判明したことにより、人工知能研究は再び冬の時代を迎える⁶⁾。

この問題を解決したのが、2006年にGeoffrey Hintonたちが提案した「オートエンコーダ（autoencoder）」である。それまでのニューラルネットワークでは結合強度の初期値をランダムに割り当てていたが、Hintonたちは、誤差逆伝播法での勾配消失を防ぐために、オートエンコーダで事前に訓練したものを初期値として用いるという事前学習を採用した。

オートエンコーダとは、入力データを再現するデータを出力するよう中間層をトレーニングするニューラルネットワークのことである。このネットワークは、入力層とそれより少ないユニット数の中間層、そして入力層と同じユニット数から成る出力層で構成される。中間層のユニット数を入力層より少なくすることにより、入力データの「特徴」を抽出することができる。この特徴は、ネットワークの結合強度を調整することによって得られる。この過程を「自己符号化」という。目標とする多層パーセプトロンの中間層に対し、1層ずつ自己符号化を進め、得られた結合強度を使って層同士をつなぎ合わせていく。このようにオートエンコーダを用いて多層パーセプトロンに事前学習をさせると、結合強度の初期値に完全なランダム値を割り当てたときと比べ、勾配消失の起こる可能性が小さくなることが判明した。これによってネットワークの深層化が可能となったため、ニューラルネットワークはディープラーニングへとさらなる進化を遂げ、現在の人工知能ブームへとつながることとなった（岡谷 2015）。

5. GB理論 vs. ニューラルネットワーク

人間の脳をコンピュータに見立てたChomskyとコンピュータで人間の脳の働きを再現しようとしたニューラルネットワークは、確かに計算主義と

いう出発点は共有していたが、その後、向かっていくベクトルの方向は正反対であった。このような方向性の違いのため、この二つのアプローチは時にぶつかり合い、論争を巻き起こすことになる。

先述の母語習得における「刺激の貧困」の問題について、Chomsky は、間接的な否定証拠のみで（つまり肯定証拠のみで）母語習得が成り立つのは、普遍文法が生得的に頭の中に備わっているからであり、それが制約としての機能を果たしているからだと主張している。これに対し Rumelhart と McClelland は、ニューラルネットワークからの反証を試みている。つまり、普遍文法による制約がなくても肯定証拠と間接的な否定証拠のみで規則の習得が可能であることを、コンピュータによるシミュレーションにより実証したのである (Rumelhart & McClelland 1986)。

彼らを取り上げたのは、規則動詞と不規則動詞の過去形を形成する英語の文法規則の習得だった。設定されたのは、原形の動詞を入力すると、それに対応する過去形が出力されるパーセプトロン型のネットワークである。トレーニングでは、正しい過去形のパターンを教師信号として、ユニット間の結合強度を調整しながら試行を繰り返していく。その際、Rumelhart たちは子どもの習得段階に倣って、まず高頻度語 10 語（このうち 8 語が不規則動詞、規則動詞は 2 語）についてのみ学習させた。すると、これら 10 語それぞれに対し 10 回ずつの試行と結合強度の調整を繰り返した段階で、ネットワークは対応する過去形をかなり良好にマスターした。そこで、第 2 段階として中頻度語 410 語（このうち不規則動詞は 76 語のみ、残りは全て規則動詞）も学習に加えた。合計 420 語の単語それぞれについて 190 回の試行と結合強度の調整を繰り返したところ、ネットワークは、これら 420 語の過去形をほとんど誤りの無いレベルまでマスターした。しかも、このネットワークは、学習時には導入されていない低頻度語 86 語（規則動詞 72 語・不規則動詞 14 語）についても 91%（規則動詞 92%，不規則動詞 84%）という高い確率で正しい過去形を出力できた。

このように Rumelhart たちは、普遍文法による制約がなくても肯定証拠と間接的な否定証拠のみ（教師信号が該当）で文法規則の習得が可能であることを示した。このことは Chomsky の主張に対するニューラルネットワークからの反証であると

言える。これに対し、Pinker & Prince (1988) は実際の子どもの母語話者による言語使用の割合について調べ、必ずしも Rumelhart たちのモデルのようにはなっていないと反論している。

Pinker たちは、幼児の英語母語話者（3 人）の主に母親との会話を 4 年間にわたって録音、収集した有名な Brown (1973) のデータから規則動詞と不規則動詞の過去形の言語使用の割合を分析した。すると、いずれの子においても規則動詞と不規則動詞の過去形の割合は、初期段階で不規則動詞に偏ることも、また、その後に規則動詞に偏ることもなく、どの段階においても規則動詞と不規則動詞とが大体半々の割合だった。

この Pinker たちによる指摘は重要だが、それ以前に実際の子どもたちを取り巻く言語環境自体が、必ずしも子どもたちの習得段階を考慮したものとはなっていないという問題がある。つまり、子どもたちは常に周りで話されている言葉を耳にしているため、自身の習得段階に関係なく、あらゆる語彙の入力にさらされているはずであり、その事実自体が先述の Rumelhart たちのシミュレーションとは矛盾している。

ニューラルネットワークの観点から、この疑問に答えるものとして Jeffrey Elman による一連の研究がある⁷⁾。Elman がシミュレーションに用いたのはパーセプトロン型のネットワークではなく、単純再帰ネットワーク（エルマンネット）と呼ばれるものだった。単純再帰ネットワークでは、入力層と出力層の間の中間層の状態を一時的に記憶する「文脈層 (context units)」を配して、語順のような時系列データの処理を可能にしている。Elman (1991) は、この単純再帰ネットワークを用いて英語の関係詞節を含んだ複文構造をコンピュータに学習させた。

しかし、Elman (1991) でも Rumelhart たちのシミュレーションと同じように段階に沿った学習を導入している。まず第 1 段階では関係詞節を含まない単文のみ 10,000 文が提示され、第 2 段階では提示された 10,000 文のうちの 25% が関係詞節を含む複文で残り 75% が単文、第 3 段階では 10,000 文のうちの半数が関係詞節を含む複文で残り半数が単文、第 4 段階では 10,000 文のうちの 75% が関係詞節を含む複文で単文は 25% であった。トレーニングでは、各段階で 10,000 文それぞれに対し 5 回ずつの試行とそれに伴う結合強度の

調整が繰り返されたので、各段階で 50,000 試行、段階全体では 200,000 試行となった。トレーニング終了後、結合強度を固定して、出力ユニットそれぞれの活性化の程度から正しい単語の生起する確率について調べたところ、ネットワークが関係詞節を正しく認識できていることが示された。

このシミュレーションでも段階に沿った学習が導入されているが、先述したように実際の子どもたちを取り巻く言語環境は必ずしも段階を考慮したものにはなっていない。そこで Elman (1993) では、学習者自身が段階的に発達していく過程に注目し、子どもの作動記憶におけるメモリスパンが発達に応じて増えていくことをネットワーク上で再現しようとした。

その際、Elman が利用したのが単純再帰ネットワークにおける文脈層の調整である。文脈層における全てのユニットの活性化値を、あるタイミングで強制的に中立的な 0.5 の値に調整すると、それまでの記憶が減衰してしまう状態を再現できる。第 1 段階では、そのタイミングを 3 または 4 語ごとにしてメモリスパンが 3 ないし 4 語分しかない状態を再現した。第 2 段階では、この調整のタイミングを 4 ないし 5 語ごとにずらし、第 3 段階では 5 ないし 6 語ごとに、第 4 段階では 6 ないし 7 語ごとにずらした。そして第 5 段階では、そのような強制的な調整は一切行わなかった。このようにしてメモリスパンが徐々に増えていく状態を再現しながらネットワークのトレーニングを行ったところ、ネットワークは最初から関係詞節を含む複雑な文例が入力されていたにもかかわらず学習を達成することができた⁸⁾。

以上のことから、シミュレーションでは文脈層の機能制限によって再現したように、母語習得においては作動記憶におけるメモリスパンが発達に応じて増えていく認知的な発達段階それ自体が制約の役割を果たしている可能性が示唆された。つまり、生得的な普遍文法ではなく認知的な発達それ自体が制約の機能を果たしている可能性が示されたのである (Elman 2001)。しかしながら、ニューラルネットワークについては学習を成立させるためのアルゴリズムについて生物学的な妥当性の問題もあるため、上記のようなシミュレーションの結果が、そのまま人間による「学び」を再現しているとは言い切れず、論争の決着にまでは至っていない。

6. 結語

Chomsky が人間の脳をコンピュータに見立てたのに対し、ニューラルネットワークではコンピュータで人間の脳の働きを再現しようとしている。そのためニューラルネットワークにおいては、コンピュータが本来の目的である「再現」を乗り越えて人間の脳の働きを超えてしまうこと (シンギュラリティ) は大問題だが、真逆の方向からアプローチしている Chomsky にとっては、

it would be of no interest to show that ...
a computer program that could 'beat' a
grandmaster in chess is about as interesting
as a bulldozer that can 'win' the Olympic
weight-lifting competition. (Chomsky 1996:
40)

(コンピュータがチェスで人間のチャンピオンを破ったことなど、ブルドーザーがオリンピックの重量挙げで優勝するのと同じで全く興味がない) 《和訳は筆者が添付》

ということになる。

結語として、筆者自身としてはニューラルネットワークによる脳の働きの再現に今後の研究の可能性を見出している。しかし、それは冒頭の松尾氏が言うように普遍文法を人工知能に埋め込むことができるかと期待しているからではない。むしろ、シミュレーションによる知見を積み重ねていくことにより、普遍文法という概念に頼らなくても、それと同様の機能を果たすことができるシステムの解明につながるのではないかと期待しているからである。

注

1) Kurzweil (2005) は、コンピュータの演算能力の観点から、2030 年代の始めには人工知能が人間の知能と同等のレベルに達し、2045 年にはシンギュラリティに至るだろうと予測している。そのためシンギュラリティは「2045 年問題」とも呼ばれている。

2) しかしながら、このようなパラメータを想定することは、結果として多くの「マイクロ・パラメータ」を生み出すこととなった。つまり、このモデルを実際に存在する言語に適用しようとした場合、すんなりと二項対立的な図式に収まるものば

かりではないので、逆にそれらに対応する新たなパラメータを設定する必要に迫られてしまう。そのためパラメータの細分化が進み、結果的に多くのマイクロ・パラメータが提案されることとなった。このような事態を懸念した Chomsky は、1980 年代の終わり頃から「経済性の原理」という考えのもと、このような余剰性の排除を主張し始めるようになり、1990 年頃からは「ミニマリスト・プログラム」という新しい枠組みを掲げて、この考えをさらに推し進め、余剰性を排したよりシンプルなモデルの提案を試みている (吉田 1997)。

3) 本節における子どもが利用可能な証拠に関する記述については Cook (1988: 59-62) を参照した。また、後述の「間接的な否定証拠」については、それを支えているが文法規則に適った言語の正しい用例であることから、ここでは肯定証拠と同義で扱っている。

4) 本節でのニューラルネットワークにおける歴史的な研究の推移については、大野 (2015) と深澤 (2015) を参照した。

5) ここでの「誤差逆伝播法」の学習アルゴリズムに関する記述は、McLeod, Plunkett & Rolls (1998) を参照した。

6) この「勾配消失」の問題に逸早く着目したのは Sepp Hochreiter である。彼は 1991 年に提出した学位論文 (ドイツ語) の中で、すでに「誤差逆伝播法」におけるこの欠点について指摘している。

7) ここでの Elman による一連の研究については、守 (2001) を参照した。

8) この際、初期において十分なトレーニングがなされないと、その後の学習に支障が出ることもわかった。つまり、第 1 段階では 10,000 文それぞれに対して 12 回ずつの試行を繰り返す必要があった。しかし、その後の段階については Elman (1991) と同じ 5 回の試行で学習が進行した。

引用文献

- Bloomfield, L. (1933) *Language*. New York: Henry Holt & Company. (三宅鴻・日野資純 (訳) (1962) 『言語』, 大修館書店.)
- Brown, R. (1973) *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.
- Chomsky, N. (1957) *Syntactic Structures*. The Hague: Mouton. (勇康雄 (訳) (1963) 『文法

の構造』, 研究社.)

Chomsky, N. (1959) A review of B. F. Skinner's *Verbal Behavior*. *Language*, 35(1), 26-58.

Chomsky, N. (1964) *Current Issues in Linguistic Theory*. The Hague: Mouton.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press. (安井稔 (訳) (1970) 『文法理論の諸相』, 研究社.)

Chomsky, N. (1981) *Lectures on Government and Binding: The Pisa Lectures*. Holland: Foris Publications. (安井稔・原口庄輔 (訳) (1986) 『統率・束縛理論』, 研究社出版.)

Chomsky, N. (1988) *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: MIT Press. (田窪行則・郡司隆男 (訳) (1989) 『言語と知識: マナグア講義録 (言語学編)』, 産業図書.)

Chomsky, N. (1996) *Powers and Prospects: Reflections on Human Nature and the Social Order*. London: Pluto Press.

Cook, V. J. (1988) *Chomsky's Universal Grammar: An Introduction*. Cambridge, MA: Basil Blackwell. (須賀哲夫 (訳) (1990) 『チョムスキーの言語理論: 普遍文法入門』新曜社.)

Elman, J. (1991) Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7: 195-225.

Elman, J. (1993) Learning and development in neural networks: The importance of starting small. *Cognition*, 48: 71-99.

Elman, J. (2001) Connectionism and language acquisition. In M. Tomasello & E. Bates, Eds. *Language Development: The Essential Readings* (pp. 295-306). Malden, MA: Blackwell Publishers.

深澤祐援 (2015) 「ディープラーニングとは何なのか? そのイメージをつかんでみる」, 『ITmedia ビジネスオンライン (ITmedia Inc.)』
<http://bizmakoto.jp/makoto/articles/1507/27/news067.html> (最終閲覧日 2017 年 9 月 23 日)

林龍平 (1984) 「学習過程」, 馬場道夫・吉岡伸 (編) 『人間教育の心理学』 (pp. 71-85), 第一法規.

Hinton, G. E. & Salakhutdinov, R. R. (2006) Reducing the dimensionality of data with

- neural networks. *Science*, 313 (5786): 504-507.
- Hochreiter, S. (1991) *Untersuchungen zu dynamischen neuronalen Netzen*. Diploma thesis, Technische Universität München.
- Kurzweil, R. (2005) *The Singularity is Near: When Humans Transcend Biology*. New York: Viking Adult. (井上健, 小野木明恵, 野中香方子, 福田実 (訳) (2007) 『ポスト・ヒューマン誕生: コンピュータが人類の知性を超えるとき』, 日本放送出版協会.)
- 松尾豊 (2015) 『人工知能は人間を超えるか: ディープラーニングの先にあるもの』, KADOKAWA.
- McLeod, P., Plunkett, K. & Rolls, E. T. (1998) *Introduction to Connectionist Modeling of Cognitive Processes*. Oxford: Oxford University Press. (深谷澄男 (監訳) (2005) 『認知過程のコネクショニスト・モデル』, 北樹出版.)
- Minsky, M. & Papert, S. (1969) *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press.
- 守一雄 (2001) 「単純再帰ネットワーク (エルマン ネット) による文法の獲得」, 守一雄・都築誉史・楠見孝 (編) 『コネクショニストモデルと心理学: 脳のシミュレーションによる心の理解』 (pp. 26-37), 北大路書房.
- 大野健太 (2015) 「ニューラルネットの歩んだ道, ディープラーニングの登場で全てが変わった」, 『日経ビッグデータ (日経 BP 社)』 <http://business.nikkeibp.co.jp/article/bigdata/20150419/280107/?P=1> (最終閲覧日 2017 年 9 月 23 日)
- 岡谷貴之 (2015) 『深層学習 (機械学習プロフェッショナルシリーズ)』, 講談社.
- Pinker, S. & Prince, A. (1988) On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28: 73-193.
- Rosenblatt, F. (1958) The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65 (6): 386-408.
- Rumelhart, D. E. & McClelland, J. L. (1986) On learning the past tenses of English verbs. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group, Eds. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 2, pp. 216-271. Cambridge, MA: MIT Press.
- Rumelhart, D. E., McClelland, J. L. & the PDP Research Group, Eds. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vols. 1 & 2. Cambridge, MA: MIT Press. (甘利俊一 (監訳) (1989) 『PDP モデル—認知科学とニューロン回路網の探索—』, 産業図書.)
- Skinner, B. F. (1957) *Verbal Behavior*. New York: Appleton-Century-Crofts.
- 吉田智行 (1997) 「生成文法理論のチャレンジ: ミニマリスト・プログラムへの展開」, 『アジア文化研究』 (国際基督教大学アジア文化研究所), 別冊 7 号, 160-178.